Robust estimates for ARMA models

Article *in* Journal of the American Statistical Association · March 1986

DOI: 10.1080/01621459.1986.10478253

2 authors:

Oscar Bustos
National University of Cordoba, Argentina

Victor J. Yohai
Universidad de Buenos Aires

Robust Estimates for ARMA Models
Author(s): Oscar H. Bustos and Victor J. Yohai
Source: *Journal of the American Statistical Association*, Vol. 81, No. 393 (Mar., 1986), pp. 155-168
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2287983
Accessed: 22-12-2015 14:05 UTC

Robust Estimates for ARMA Models

OSCAR H. BUSTOS and VICTOR J. YOHAI*

Two new classes of robust estimates for ARMA models are introduced: estimates based on residual autocovariances (RA estimates), and estimates based on truncated residual autocovariances (TRA estimates). A heuristic derivation of the asymptotic normal distribution is given. We also perform a Monte Carlo study to compare the robustness properties of these estimates with the least squares, M, and GM estimates. In this study we consider observations that correspond to a Gaussian model with additive outliers. The Monte Carlo results show that RA and TRA estimates compare favorably with respect to least squares, M, and GM estimates.

KEY WORDS: Robust estimation; Monte Carlo.

1. INTRODUCTION

Suppose that z_t $(1 \leq t \leq T)$ are observations corresponding to a stationary and invertible autoregressive moving average (ARMA) process, with p autoregressive parameters and q moving average parameters (ARMA(p, q)). Then we have

$$\boldsymbol{\phi}(B)(z_t - \mu) = \boldsymbol{\theta}(B)u_t, \tag{1.1}$$

where u_t are independent identically distributed random variables, μ is the mean of z_t, the operators $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are given by

$$\boldsymbol{\phi}(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \tag{1.2}$$

and

$$\boldsymbol{\theta}(B) = 1 - \theta_1 B - \cdots - \theta_q B^q, \tag{1.3}$$

and B is the backward shift operator defined by $Bz_t = z_{t-1}$.

Box and Jenkins (BJ, 1970) proposed estimates for $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$, and μ (BJ estimates) that are asymptotically equivalent to the maximum likelihood estimates. Let $\boldsymbol{\lambda} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$. Then the BJ estimates are defined as the values of $\boldsymbol{\lambda}$ minimizing

$$\sum_{t=-\infty}^{T} E^2(u_t \mid \mathbf{z}, \boldsymbol{\lambda}), \tag{1.4}$$

where $\mathbf{z} = (z_1, \ldots, z_T)'$ and $E(u_t \mid \mathbf{z}, \boldsymbol{\lambda})$ denotes the conditional expectation of u_t given $\mathbf{z}$, when the parameters are $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, μ.

On the other hand, the least squares (LS) estimates of $\boldsymbol{\lambda}$ minimize

$$\sum_{t=p+1}^{T} r_t^2(\boldsymbol{\lambda}), \tag{1.5}$$

where the residuals $r_t(\boldsymbol{\lambda})$ are defined by

$$r_t(\boldsymbol{\lambda}) = \boldsymbol{\theta}(B)^{-1}\boldsymbol{\phi}(B)(z_t - \mu), \tag{1.6}$$

and it is assumed that $z_t = \mu$ for $t \leq 0$. We will write r_t for the residuals instead of $r_t(\boldsymbol{\lambda})$ when this does not cause confusion. The residuals may be computed recursively from

$$r_t = (z_t - \mu) - \phi_1(z_{t-1} - \mu) - \cdots - \phi_p(z_{t-p} - \mu) + \theta_1 r_{t-1} + \cdots + \theta_q r_{t-q}, \tag{1.7}$$

for $t \geq p + 1$, with initial conditions $r_p = \cdots = r_{p-q+1} = 0$.

It is well known (see Box and Jenkins 1970) that the BJ and LS estimates are asymptotically equivalent and efficient when the u_t have a normal distribution. Neither method is robust, however, since they are both very sensitive to a few outliers or abnormal observations. Moreover, they may be inefficient when the distribution of u_t has heavy tails.

The goals of robust estimation are to find estimates (a) that are highly efficient under a central model and (b) such that "small" changes in the distribution of the sample produce "small" changes in the distribution of the estimates. Another concept (introduced by Tukey 1976), closely related to robustness and which may be considered its data-oriented counterpart, is resistance. A resistant sequence of estimates should be (a) stable in the presence of a few atypical outliers (i.e., a small percentage of abnormal observations should not influence the estimates too much) and (b) stable when all of the observations have small errors (e.g., round-off errors).

Formal definitions of qualitative robustness of estimates when the observations correspond to a stochastic process have been given by Papantoni-Kazakos and Gray (1979), Bustos (1981), Cox (1981), and Boente, Fraiman, and Yohai (1982). These definitions are generalizations of Hampel's (1971) definition for the iid case, and they establish that a sequence of estimates is robust if a small change in the distribution of the process produces a small change in the distribution of the estimate uniformly on the sample size T. A different approach to qualitative robustness is also given in Boente et al. (1982).

They define strong pointwise robustness, formalizing the intuitive meaning of resistance as insensitivity to a small percentage of gross errors and to small errors in all of the observations.

To study the robustness of the LS estimates and the other proposals, we consider the two types of outliers studied in Denby and Martin (1979): (a) innovation outliers and (b) additive outliers.

Innovation Outliers. We say that z_t is an ARMA(p, q) time series with innovation outliers if z_t satisfies (1.1) but the u_t's have a heavy-tailed nearly normal distribution F—for example, F is normal contaminated, given by

$$F = (1 - \varepsilon)N(0, \sigma^2) + \varepsilon G, \tag{1.8}$$

where ε is small, $N(\mu, \sigma^2)$ stands for a normal distribution with mean μ and variance σ^2, and G is an arbitrary distribution with

* Oscar H. Bustos is Assistant Researcher, Instituto de Matemática Pura e Aplicada (IMPA), Rua Dona Castorina, 110, 22460, Jardim Botânico, Rio de Janeiro, Brazil. Victor J. Yohai is Professor, Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 1, 1428 Buenos Aires, Argentina, and Senior Researcher, Centro de Estudios de Macroeconomia de Argentina, Virrey del Pino 3210, Buenos Aires, Argentina. The authors thank Elena Martinez and Elena Ganon for undertaking some of the computations, R. D. Martin, and the referees for very helpful comments and suggestions.

Journal of the American Statistical Association
March 1986, Vol. 81, No. 393, Theory and Methods

a dispersion greater than σ [for example, $G = N(0, \tau^2)$ with $\tau^2 \geq \sigma^2$].

This means that the innovations u_t come from an $N(0, \sigma^2)$ with probability $1 - \varepsilon$ and from an arbitrary distribution G having greater dispersion with probability ε. The u_t's coming from G may be considered outliers.

The important feature of this kind of outlier is that the dynamic equation (1.1) is always satisfied, even if the u_t's in the right-hand side are outliers. In this case the process z_t is a perfectly observed ARMA(p, q) process.

Additive Outliers. In this case the observed z_t series is not itself an ARMA process. Instead, we have

$$z_t = w_t + v_t, \tag{1.9}$$

where w_t is an ARMA process satisfying (1.1) with $u_t \sim N(0, \sigma^2)$ and v_t is an independent sequence of variables, independent of the sequence w_t. The variables v_t have distribution H, given by

$$H = (1 - \varepsilon)\delta_0 + \varepsilon G, \tag{1.10}$$

where δ_0 is the distribution that assigns probability 1 to the origin and G is an arbitrary distribution. Therefore, with probability $1 - \varepsilon$, the ARMA(p, q) process w_t itself is observed, and with probability ε the observation is the ARMA(p, q) process w_t plus an error with distribution G.

The LS estimate is not very sensitive to the presence of innovation outliers. It may be shown (see Box and Jenkins 1970 and Whittle 1962) that if the innovations have finite variance, the asymptotic covariance matrix of $T^{1/2}(\hat{\boldsymbol{\phi}}_{\text{LS}} - \boldsymbol{\phi}, \hat{\boldsymbol{\theta}}_{\text{LS}} - \boldsymbol{\theta})$ is independent of the distribution F. Therefore the presence of innovation outliers does not inflate the variance of $\hat{\boldsymbol{\phi}}_{\text{LS}}$ and $\hat{\boldsymbol{\theta}}_{\text{LS}}$. On the contrary, the variance of $T^{1/2}(\hat{\mu}_{\text{LS}} - \mu)$, which depends on $E_F(u_t^2)$, can be tremendously inflated under the presence of innovation outliers.

Yohai and Maronna (1978) have proved that the LS estimates are consistent for the parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ of a stationary autoregressive process with innovations u_t with symmetric distribution F, even in the case when $E_F(|u_t|) = \infty$. Kanter and Steiger (1974) have shown that the order of convergence of the LS estimates may be even faster than $T^{1/2}$ in the case when F is a symmetric stable law with infinite variance.

The LS estimate, however, may be highly inefficient under innovation outliers compared with other estimates—for example, the corresponding maximum likelihood estimator (see Denby and Martin 1979). A class of robust estimates for ARMA models with innovation outliers is given by the class of maximum-likelihood-type estimates (M estimates) (for example, see Lee and Martin 1982).

On the other hand, the LS estimate and even the M estimates are extremely sensitive to the presence of additive outliers. This fact is reported in Denby and Martin (1979) for autoregressive processes. In Section 4 we report the results of a Monte Carlo simulation, which show that for an MA(1) model, the LS estimate is even more sensitive to additive outliers than in the case of autoregressions. Since a small fraction of additive outliers may cause a large bias in the LS and M estimates, they are much more dangerous than innovation outliers, and the only way to control them is through robust estimation. It is for this reason that we are especially interested in finding estimates that are robust under additive outliers.

Two classes of estimates that are robust in the presence of additive outliers have been proposed: (a) general maximum-likelihood-type estimates (GM estimates) for autoregressive processes (see Denby and Martin 1979, Martin 1980, and Bustos 1982) and (b) approximate maximum likelihood estimates (AM estimates) for ARMA models (see Martin 1981). The GM estimates are consistent under a perfectly observed autoregressive model, but they do not completely use the structure of time series when downweighting observations. Consequently, they have a complicated asymptotic covariance matrix, and the calibration of the tuning constant when making robust the function that defines the estimates depends on the order of the autoregressive operator. The AM estimators seem to behave quite well; even though the asymptotic biases may be small (see Martin 1981 for evidence along these lines), they are not consistent under the perfectly observed ARMA model.

In this article we present two new families of estimates that are robust and resistant in the presence of additive outliers: estimates based on residual autocovariances (RA estimates) and estimates based on truncated residual autocovariances (TRA estimates). Both classes, the RA and TRA estimates, are Fisher-consistent for any stationary and invertible ARMA model with symmetrically distributed innovations.

The RA estimates seem to be qualitatively robust only for stationary autoregressive models. Monte Carlo results show, however, that they are more stable than LS and M estimates for moving average models with additive outliers. These estimates have the attractive feature that their asymptotic covariance matrix is very simple: it is the same as that of the LS estimate, except for a scalar factor that is independent of the ARMA model. This greatly simplifies the problem of adjusting the tuning constants used for the estimates. We give an interpretation of a subclass of the RA estimates as iterated LS estimates of a Winsorized process.

The TRA estimates seem to be qualitatively robust for any stationary and invertible ARMA model.

In Section 2 we introduce the RA estimates. In Section 3 we introduce the TRA estimates. In Section 4 we show the results of a Monte Carlo study comparing the TRA, RA, GM, M, and LS estimates for the AR(1) and MA(1) models. In Appendix A we give a heuristic derivation of the asymptotic distributions of the RA and TRA estimates. In Appendix B we give some details concerning random number generation and the numerical algorithms used for computing the estimates.

2. ESTIMATES BASED ON RESIDUAL AUTOCOVARIANCES

2.1 Definition of the RA Estimates

The basic idea behind the RA estimates is to exhibit the least squares estimates in a form that involves (the usual nonrobust) covariance estimates of residuals, and then make LS robust by making the covariance estimates robust in a natural way.

Consider the LS estimates obtained minimizing (1.5). Differentiating this expression we obtain the following system of

$p + q + 1$ equations for the LS estimates:

$$\sum_{t=p+1}^{T} r_t(\partial r_t/\partial\phi_j) = 0, \qquad 1 \le j \le p$$

$$\sum_{t=p+1}^{T} r_t(\partial r_t/\partial\theta_j) = 0, \qquad 1 \le j \le q$$

$$\sum_{t=p+1}^{T} r_t(\partial r_t/\partial\mu) = 0. \tag{2.1}$$

It is easy to show from (1.1) that

$$\partial r_t/\partial\phi_j = -\boldsymbol{\theta}^{-1}(B)B^j(z_t - \mu) = -\boldsymbol{\phi}^{-1}(B)r_{t-j}$$

$$\partial r_t/\partial\theta_j = \boldsymbol{\theta}^{-2}(B)\boldsymbol{\phi}(B)B^j(z_t - \mu) = \boldsymbol{\theta}^{-1}(B)r_{t-j}$$

$$\partial r_t/\partial\mu = -(1 - \phi_1 - \cdots - \phi_p)/(1 - \theta_1 - \cdots - \theta_q), \tag{2.2}$$

where stationarity and invertibility of the ARMA process ensures that $\boldsymbol{\phi}^{-1}(B)$ and $\boldsymbol{\theta}^{-1}(B)$ exist.

By replacing parameter values with estimated parameter values in (2.2) and using the result in (2.1), we obtain that the LS estimates satisfy

$$\sum_{t=p+1}^{T} r_t\boldsymbol{\phi}^{-1}(B)r_{t-j} = 0, \qquad 1 \le j \le p$$

$$\sum_{t=p+1}^{T} r_t\boldsymbol{\theta}^{-1}(B)r_{t-j} = 0, \qquad 1 \le j \le q$$

$$\sum_{t=p+1}^{T} r_t = 0. \tag{2.3}$$

Call $s_i = s_i(\boldsymbol{\phi})$, $0 \le i < \infty$, the series expansion coefficients of the operator $\boldsymbol{\phi}^{-1}(B)$ and $t_i = t_i(\boldsymbol{\theta})$, $0 \le i < \infty$, those of $\boldsymbol{\theta}^{-1}(B)$; that is,

$$\boldsymbol{\theta}^{-1}(B) = \sum_{i=0}^{\infty} s_iB^i, \quad \boldsymbol{\theta}^{-1}(B) = \sum_{i=0}^{\infty} t_iB^i. \tag{2.4}$$

Then we can write (2.3) as

$$\sum_{t=p+1+j}^{T} r_t \sum_{h=0}^{\infty} r_{t-j-h}s_h = 0, \qquad 1 \le j \le p$$

$$\sum_{t=p+1+j}^{T} r_t \sum_{h=0}^{\infty} r_{t-j-h}t_h = 0, \qquad 1 \le j \le q$$

$$\sum_{t=p+1}^{T} r_t = 0.$$

Using the initial conditions $r_t = 0$ for $t < p + 1$ and interchanging the order of summations, we obtain

$$\sum_{h=0}^{T-j-p-1} s_h\gamma_{h+j}(\boldsymbol{\lambda}) = 0, \qquad 1 \le j \le p$$

$$\sum_{h=0}^{T-j-p-1} t_h\gamma_{h+j}(\boldsymbol{\lambda}) = 0, \qquad 1 \le j \le q$$

$$\sum_{t=p+1}^{T} r_t = 0, \tag{2.5}$$

where

$$\gamma_i(\boldsymbol{\lambda}) = \sum_{t=p+1}^{T-i} r_{t+i}r_t. \tag{2.6}$$

We now define the class of robust RA estimates by making the residual covariances γ_i robust. This may be done by replacing the $\gamma_i(\boldsymbol{\lambda})$ in (2.5) by

$$\gamma_i(\boldsymbol{\lambda}) = \sum_{t=p+1+i}^{T} \eta(r_t/\hat{\sigma}, r_{t-i}/\hat{\sigma}), \qquad i = 1, 2, \ldots. \tag{2.7}$$

In the last equation of (2.5), r_t is replaced by $\psi(r_t/\hat{\sigma})$, where $\eta: R^2 \to R$ and $\psi: R \to R$ are bounded and continuous functions and $\hat{\sigma}$ is a robust estimate of the scale of the u_t's. Then the RA estimates are defined by the following $p + q + 1$ equations:

$$\sum_{h=0}^{T-j-p-1} s_h\gamma_{h+j}(\boldsymbol{\lambda}) = 0, \qquad 1 \le j \le p$$

$$\sum_{h=0}^{T-j-p-1} t_h\gamma_{h+j}(\boldsymbol{\lambda}) = 0, \qquad 1 \le j \le q$$

$$\sum_{t=p+1}^{T} \psi(r_t/\hat{\sigma}) = 0. \tag{2.8}$$

$\hat{\sigma}$ is computed simultaneously using, for example,

$$\hat{\sigma} = \text{Med}(|r_{p+1}|, \ldots, |r_T|)/.6745. \tag{2.8$'$}$$

We will assume throughout the article that η is odd in each variable and ψ is odd. Therefore, if the distribution of the innovations u_t is symmetric, the RA estimates will be Fisher-consistent. This is because in this case we have

$$\lim_{t\to\infty} r_t(\boldsymbol{\lambda}_0) = u_t,$$

where $\boldsymbol{\lambda}_0$ is the true parameter and

$$E_F(\eta(u_t/\sigma, u_{t-i}/\sigma)) = 0 \;\forall\; \sigma, \qquad i = 1, 2, \ldots$$

$$E_F(\psi(u_t/\sigma)) = 0 \;\forall\; \sigma. \tag{2.9}$$

Two ways of choosing η are

$$\eta_M(u, v) = \psi(u)\psi(v), \qquad \text{Mallows type}$$

$$\eta_H(u, v) = \psi(uv), \qquad \text{Hampel type,} \tag{2.9$'$}$$

where ψ is a continuous and odd function. They may be chosen, for example, to be in the Huber family given by

$$\psi_{H,c}(u) = \text{sgn}(u)\min(|u|, c), \tag{2.10}$$

where $\text{sgn}(u)$ is the sign function.

Another possibility is to take the ψ function in a redescending family—for example, in the bisquare family proposed by Beaton and Tukey (1974). This family is defined by

$$\psi_{B,c}(u) = u(1 - u^2/c^2)^2, \qquad 0 \le |u| \le c. \tag{2.11}$$

Remark 1. If $\eta(u, v) = \psi(u)v$, then the RA estimate given by (2.8) is asymptotically equivalent to the M estimate minimizing

$$\sum_{t=p+1}^{\infty} \rho(r_t/\sigma),$$

where $\rho' = \psi$. Therefore the class of RA estimates contains the class of M estimates. In particular the LS estimate is obtained putting $\eta(u, v) = uv$ and $\psi(u) = u$.

Remark 2. Since the s_h and t_h decay exponentially, for numerical computations it is not necessary to consider all of the terms in the first $p + q$ equations of (2.8).

2.2 Computing RA Estimates

A heuristically appealing interpretation of the RA estimates may be obtained in the case $\eta(u, v) = \psi(u)\psi(v)$. Put

$$r_t^* = \psi(r_t(\hat{\boldsymbol{\lambda}})/\hat{\sigma})\hat{\sigma} \tag{2.12}$$

and

$$z_t^* = \hat{\mu} + \hat{\boldsymbol{\phi}}^{-1}(B)\hat{\boldsymbol{\theta}}(B)r_t^*. \tag{2.13}$$

If $\psi = \psi_{H,c}$, given by (2.10), the series z_t^* may be regarded as a Winsorization of the z_t series. The system of Equations (2.8) implies that its solution $\hat{\boldsymbol{\lambda}}$ is the LS estimate when the series is given by z_t^*. This property suggests the following iterative computational algorithm.

Start with initial estimates $\hat{\boldsymbol{\lambda}}^{(0)}$ and $\hat{\sigma}$. Given the estimate $\hat{\boldsymbol{\lambda}}^{(n)}$ corresponding to the nth iteration, we define $z_t^{*(n)}$ by (2.12) and (2.13) with $\hat{\boldsymbol{\lambda}}^{(n)}$ instead of $\hat{\boldsymbol{\lambda}}$. Then $\hat{\boldsymbol{\lambda}}^{(n+1)}$ is the LS estimate of the series $z_t^{*(n)}$. Simultaneous iteration on $\hat{\boldsymbol{\lambda}}$ and $\hat{\sigma}$ is also possible and desirable.

We cannot guarantee the existence of a solution of the equations defining the RA estimates (the unicity of the solution may not be guaranteed even for the LS estimate). Therefore it will be important when using an iterative algorithm to start from a reasonable estimate. This remark will be especially important when using redescending ψ functions in the definition of η. In the Monte Carlo results reported in Section 4, these issues did not cause difficulty.

2.3 Examples of RA Estimates

Consider an MA(1) model. Then the RA estimates of the parameters θ and μ are given by the solution of

$$\sum_{h=0}^{T-1} \theta^h \gamma_{h+1}(\theta) = 0$$

$$\sum_{t=2}^{T} \psi(r_t/\sigma) = 0 \tag{2.14}$$

and (2.8′), where γ_i is given by (2.7) and

$$r_t = (z_t - \mu) + \theta(z_{t-1} - \mu) + \cdots + \theta^{t-1}(z_1 - \mu). \tag{2.15}$$

Now consider an AR(1) model. Then the RA estimates of φ and μ are given by

$$\sum_{h=0}^{T-2} \phi^h \gamma_{h+1}(\phi) = 0$$

$$\sum_{t=2}^{T} \psi(r_t/\sigma) = 0 \tag{2.16}$$

and (2.8′), where

$$r_t = z_t - \mu - \phi(z_{t-1} - \mu). \tag{2.17}$$

2.4 Asymptotic Distribution of the RA Estimates

Let z_t be a stationary and invertible ARMA(p, q) process given by (1.1), and assume that the distribution F of the innovations u_t is symmetric. In Section 5 we heuristically derive the asymptotic distribution of the RA estimate defined by (2.8) and (2.9). A formal proof of the consistency and asymptotic normality of the RA estimates may be found in Bustos, Fraiman, and Yohai (1984).

Suppose $\hat{\sigma}$ converges to σ in probability. Let $\hat{\boldsymbol{\lambda}}$ be the solution of (2.8), and let $\boldsymbol{\lambda}_0$ be the true vector parameter. Then our argument of Section 5 shows that under certain reasonable conditions, $T^{1/2}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0)$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix

$$\mathbf{D}^* = \begin{bmatrix} v\mathbf{C}^{-1} & \mathbf{0}_{p+q} \\ \mathbf{0}'_{p+q} & d \end{bmatrix}, \tag{2.18}$$

where $\mathbf{C}$ is a $(p + q) \times (p + q)$ symmetric matrix given by

$$C_{i,j} = \sum_{k=0}^{\infty} s_k s_{k+j-i} \quad \text{if } i \le j \le p$$

$$C_{i,p+j} = \sum_{k=0}^{\infty} t_k s_{k+j-i} \quad \text{if } i \le p,\ j \le q,\ i \le j$$

$$C_{i,p+j} = \sum_{k=0}^{\infty} s_k t_{k+i-j} \quad \text{if } i \le p,\ j \le q,\ j \le i$$

$$C_{p+i,p+j} = \sum_{k=0}^{\infty} t_k t_{k+j-i} \quad \text{if } i \le j \le q. \tag{2.19}$$

The scalar v is given by

$$v = \sigma^2 a/b^2, \tag{2.19′}$$

where

$$a = E_F(\eta^2(u_t/\sigma, u_{t-1}/\sigma)) \tag{2.20}$$

and

$$b = E_F(\eta_1(u_t/\sigma, u_{t-1}/\sigma)u_{t-1}) \tag{2.21}$$

with

$$\eta_1(u, v) = \partial\eta(u, v)/\partial u.$$

The scalar d (the variance of $\hat{\mu}$) is given by

$$d = \sigma^2 a^*/(Hb^*)^2, \tag{2.22}$$

where

$$a^* = E_F(\psi^2(u_t/\sigma)), \tag{2.23}$$

$$b^* = E_F(\psi'(u_t/\sigma)), \tag{2.24}$$

and

$$H = (1 - \phi_1 - \cdots - \phi_p)/(1 - \theta_1 - \cdots - \theta_q). \tag{2.25}$$

As particular cases, (2.18) gives the asymptotic distribution of the M estimates when $\eta(u, v) = \psi(u)u$ and the asymptotic distribution of the LS estimate when $\eta(u, v) = uv$, $\psi(u) = u$.

For the estimation of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, the efficiency of the RA estimate with respect to the LS estimate is given by $1/v$ and for the estimation of μ by $E_F(u_t^2)b^{*2}/(a^*\sigma^2)$.

2.5 Qualitative Robustness and Resistance of the RA Estimates

We conjecture that the RA estimates are qualitatively robust and resistant for autoregressive processes in the sense of strong pointwise robustness defined in Boente et al. (1982), provided η and ψ are bounded and continuous. We do not have a formal proof of this fact, since this would require that we specify which solution of (2.8) we are considering if there is more than one. The only realistic way to do this would be by specifying the numerical algorithm used to compute the solution.

The RA estimates are not qualitatively robust when $q > 0$, since in that case a bad observation z_t will spoil all of the residuals, $u_{t'}$ with $t' > t$. The Monte Carlo results of Section 4 show, however, that even when there is a moving average component, the RA estimates are much more stable than LS or M estimates in the presence of additive outliers. Furthermore, in the next section we modify the RA estimates so that they are robust when moving average components are present.

3. ESTIMATES BASED ON TRUNCATED RESIDUAL AUTOCOVARIANCE

One way of making the RA estimates robust when moving average components are present is to replace the residual $r_t(\theta)$ by "truncated" versions. For example, in an MA(1) model with $\mu = E(z_t) = 0$ we have

$$r_t(\theta) = z_t + \theta z_{t-1} + \cdots + \theta^{t-1} z_1.$$

In this special case, given $k > 0$, the k-truncated version of the residual $r_t(\theta)$ is defined by

$$r_{t,k}(\theta) = z_t + \theta z_{t-1} + \cdots + \theta^k z_{t-k}. \tag{3.1}$$

It is immediate that if θ_0 is the true parameter, then

$$r_{t,k}(\theta_0) = u_t - \theta_0^{k-1} u_{t-k-1}.$$

Then we have the following property: $r_{t,k}(\theta_0)$ is independent of $r_{t-j,k}(\theta_0)$ for all $j > 0$, $j \neq k + 1$.

The RA estimate for this model was defined by a solution of

$$\sum_{j=1}^{T-1} \gamma_j(\theta)\theta^{j-1} = 0, \tag{3.1$'$}$$

where $\gamma_j(\theta)$ is given by (2.7). The estimates based on truncated residual autocovariance (TRA estimates) are defined similarly but by replacing $\gamma_j(\theta)$ by $\gamma_{j,k}(\theta)$ for all $j \neq k + 1$ and $\gamma_{k+1}(\theta)$ by $\gamma_{k+1,k-1}(\theta)$, where

$$\gamma_{j,k}(\theta) = \sum_{t=j+1}^{T} \eta(r_{t,k}(\theta)/\hat{\sigma}_k,\ r_{t-j,k}(\theta)/\hat{\sigma}_k) \tag{3.2}$$

and

$$\hat{\sigma}_k = \mathrm{Med}(|r_{1,k}|, \ldots, |r_{T,k}|)/.6745. \tag{3.3}$$

Assume that u_t are symmetric and θ_0 the true parameter. Then

$$E(\eta(r_{t,k}(\theta_0)/\sigma,\ r_{t-j,k}(\theta_0)/\sigma)) = 0,\ \forall j > 0,\ \forall j \neq k + 1,\ \forall \sigma$$

and

$$E(\eta(r_{t,k-1}(\theta_0)/\sigma,\ r_{t-k-1,k-1}(\theta_0)/\sigma)) = 0,\ \forall \sigma.$$

Therefore the TRA estimates satisfy the Fisher consistency condition.

Moreover, since the truncated residuals depend on only a finite number of observations, a sufficient condition for the robustness of the MA(1) TRA estimates is that η be bounded. This is also a feature of the general ARMA-model TRA estimates defined below. If k increases, the corresponding TRA estimate will be more efficient under the theoretical model without additive outliers (AO's) but less robust. Therefore the choice of k will depend on a trade-off between efficiency under the model and bias control under AO contamination.

Now we define the TRA estimates for a general ARMA (p, q) process z_t given by (1.1). In that case for any $\boldsymbol{\lambda} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$, we define the k-truncated residuals by

$$r_{t,k}(\boldsymbol{\lambda}) = \sum_{i=0}^{k} t_i w_{t-i}(\boldsymbol{\lambda}), \tag{3.4}$$

where the t_i's are defined by (2.4) and

$$\begin{aligned} w_t(\boldsymbol{\lambda}) &= \boldsymbol{\phi}(B)(z_t - \mu) \\ &= (z_t - \mu) - \phi_1(z_{t-1} - \mu) - \cdots - \phi_p(z_{t-p} - \mu). \end{aligned} \tag{3.5}$$

Since

$$r_t(\boldsymbol{\lambda}) = \sum_{i=0}^{t-1} t_i w_{t-i}(\boldsymbol{\lambda}),$$

the $r_{t,k}(\boldsymbol{\lambda})$ are truncated versions of $r_t(\boldsymbol{\lambda})$ that have the key property that they depend upon a finite number of observations z_t, uniformly in t. We will write just $r_{t,k}$ when this does not produce ambiguity. Put

$$\boldsymbol{\theta}_k^{-1}(B) = \sum_{i=0}^{k} t_i B^i;$$

therefore, we have

$$\begin{aligned} r_{t,k} &= \boldsymbol{\theta}_k^{-1}(B)\boldsymbol{\theta}(B)\boldsymbol{\theta}^{-1}(B)\boldsymbol{\phi}(B)(z_t - \mu) \\ &= \boldsymbol{\theta}_k^{-1}(B)\boldsymbol{\theta}(B)u_t = \mathbf{M}_k(B)r_t, \end{aligned}$$

where $\mathbf{M}_k(B) = \boldsymbol{\theta}_k^{-1}(B)\boldsymbol{\theta}(B)$ is a polynomial of order $q + k$ with coefficients m_k. Moreover, it is easy to verify that the first $k + 1$ coefficients of $\mathbf{M}_k(B)$ are the same as those of the identity operator—namely, $m_0 = 1$, $m_1 = m_2 = \cdots = m_k = 0$. Thus if we relabel $a_i = m_{k+i}$ $(i = 1, \ldots, q)$, then

$$r_{t,k} = r_t + a_1 r_{t-k-1} + \cdots + a_q r_{t-k-q}. \tag{3.6}$$

We will always take $k \geq q$. If $\boldsymbol{\lambda}_0$ is the true vector parameter, we have $r_t = u_t$ and then

$$\begin{aligned} &r_{t,k-q+j} \text{ is independent of } r_{t-j,k}, && j = 1, \ldots, q - 1 \\ &r_{t,k} \text{ is independent of } r_{t-j,k}, && j = q, \ldots, k \\ &r_{t,j-q-1} \text{ is independent of } r_{t-j,k}, && j = k + 1, \ldots, k + q \\ &r_{t,k} \text{ is independent of } r_{t-j,k}, && j \geq k + q + 1. \end{aligned} \tag{3.7}$$

Define $h(j)$ by

$$\begin{aligned} h(j) &= k && \text{if } q \leq j \leq k \text{ or } j \geq k + q + 1 \\ &= j - q - 1 && \text{if } k + 1 \leq j \leq k + q \\ &= k - q + j && \text{if } j \leq q - 1. \end{aligned} \tag{3.8}$$

Table 1. Constants for the Psi Functions

AREG	MH	MB	RAHH	RAMH	RAHB	RAMB
.95	1.345	4.685	2.52	1.65	9.36	5.58
.90	.990	3.900	1.79	1.34	7.66	4.65
.80	.530	3.140	1.08	.95	5.70	3.82
.70	.190	2.695	.66	.67	4.50	3.35
.60	.001	2.364	.36	.44	3.73	3.01

(Columns MH through RAMB are grouped under the spanning header "Estimates".)

Observe that h depends also on k, which is assumed fixed. From (3.7), if $\sigma > 0$ and $\sigma' > 0$ are fixed and $\boldsymbol{\lambda}_0$ is the true parameter vector, then

$$E(\eta(r_{t,h(j)}(\boldsymbol{\lambda}_0)/\sigma,\ r_{t-j,k}(\boldsymbol{\lambda}_0)/\sigma')) = 0. \tag{3.9}$$

Thus TRA estimates are Fisher-consistent when the process z_t follows a perfectly observed ARMA model—that is, $v_t \equiv 0$, provided u_t have a symmetric distribution (if η is unbounded, the u_t must have expectations, whereas for bounded η this requirement may be dropped).

Therefore, by analogy to (2.8) we define the TRA estimates by the system of equations

$$\mathbf{L}^*(\boldsymbol{\lambda}) = 0, \tag{3.10}$$

Table 2. Monte Carlo Results for the AR(1) Process

Estimates	Mean ($\varepsilon = 0$)	MSE ($\varepsilon = 0$)	Eff ($\varepsilon = 0$)	Mean ($\varepsilon = .1, \tau^2 = 9 \text{ var } w_t$)	MSE	Eff	Bias 2	RELB
				$\phi = .5$				
LS	.488	.85	1.00	.261	7.23	1.00	.057	.48
MH95	.488	.91	.93	.279	6.32	1.14	.049	.44
RAHH95	.489	.91	.93	.351	3.84	1.88	.022	.30
RAMH95	.489	.92	.92	.354	3.55	2.04	.021	.29
GMHH95	.488	.92	.92	.346	3.93	1.84	.024	.31
GMMH95	.489	.91	.93	.349	3.60	2.01	.023	.30
MB95	.488	.91	.93	.286	6.15	1.18	.046	.43
RAHB95	.490	.91	.93	.390	3.08	2.35	.012	.22
RAMB95	.489	.93	.91	.396	2.59	2.79	.011	.21
GMHB95	.489	.93	.91	.383	3.19	2.27	.014	.23
GMMB95	.489	.92	.92	.386	2.60	2.78	.013	.23
				$\phi = .8$				
LS	.782	.51	1.00	.422	16.80	1.00	.143	.47
MH95	.782	.54	.94	.507	11.40	1.47	.086	.37
RAHH95	.783	.53	.96	.647	4.10	4.10	.023	.19
RAMH95	.782	.54	.94	.637	4.35	3.86	.027	.20
GMHH95	.783	.55	.93	.616	5.21	3.23	.034	.23
GMMH95	.782	.54	.94	.609	5.28	3.18	.036	.24
MB95	.782	.54	.94	.583	8.46	1.99	.047	.27
RAHB95	.783	.53	.96	.722	1.95	8.62	.006	.10
RAMB95	.783	.54	.94	.703	2.23	7.54	.009	.12
GMHB95	.783	.55	.93	.696	2.78	6.05	.011	.13
GMMB95	.783	.54	.94	.670	2.99	5.62	.017	.16

NOTE: Sample size—100; 500 replications. See Appendix C for an explanation of the notation.

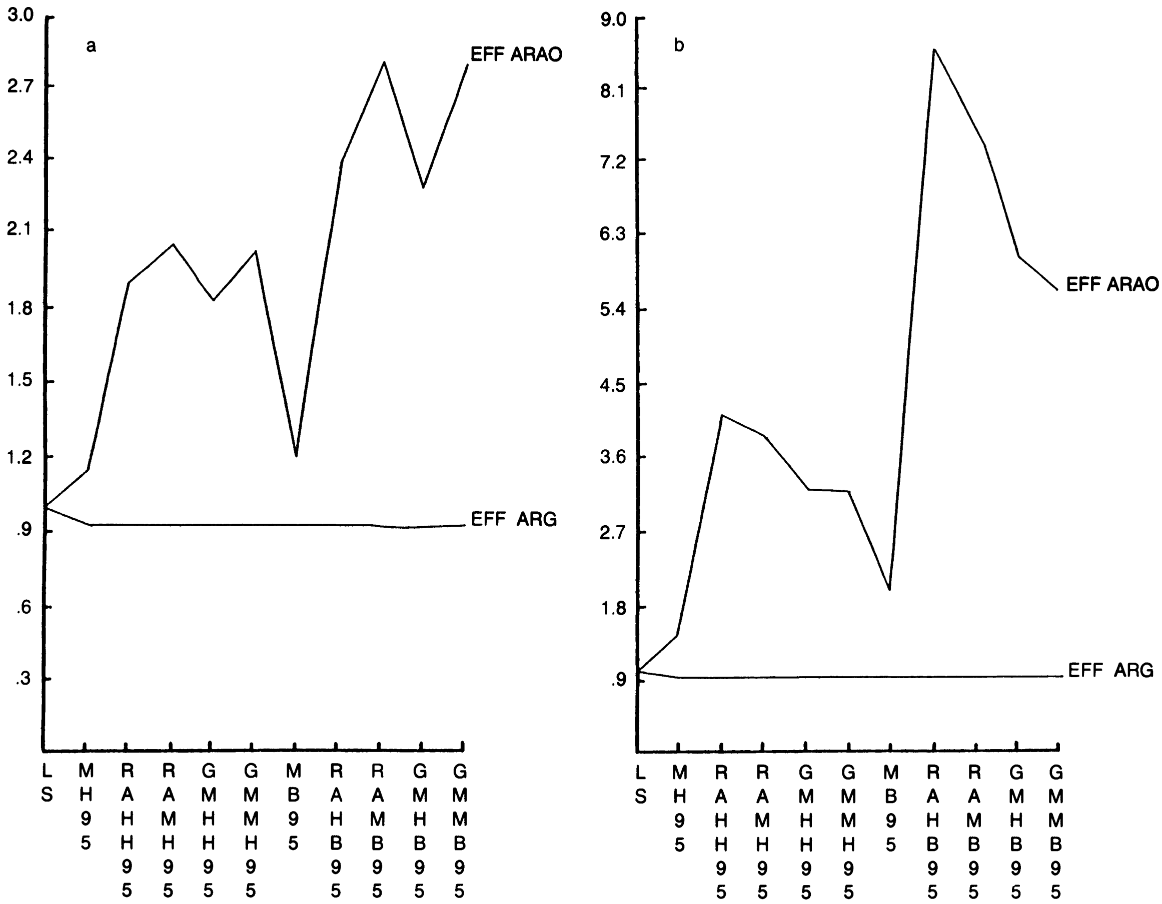

Figure 1. Efficiency of the Estimates for the ARG and ARAO Models: sample size 100; 500 replications. a—ϕ = .5; b—ϕ = .8.

Table 3. Asymptotic Results for the ARAO Model: $\varepsilon = .1$, $\tau^2 = 9 \text{ var } w_t$

	$\phi = .5$			$\phi = .8$		
Estimates	AV	Bias 2	RELB	AV	Bias 2	RELB
LS	.26	.058	.48	.42	.144	.48
*MH*95	.28	.048	.44	.53	.073	.34
RA*HH*95	.36	.020	.28	.68	.014	.15
RA*MH*95	.36	.020	.28	.66	.020	.18
GM*HH*95	.36	.020	.28	.65	.023	.19
GM*MH*95	.36	.020	.28	.63	.029	.21
*MB*95	.28	.048	.44	.64	.026	.20
RA*HB*95	.41	.008	.18	.75	.002	.06
RA*MB*95	.40	.010	.20	.72	.006	.10
GM*HB*95	.40	.010	.20	.73	.005	.09
GM*MB*95	.40	.010	.20	.69	.012	.14

NOTE: See Appendix C for an explanation of the notation.

where $\mathbf{L}^*(\boldsymbol{\lambda}) = (L_1^*(\boldsymbol{\lambda}), \ldots, L_{p+q+1}^*(\boldsymbol{\lambda}))$ is defined by

$$L_j^*(\boldsymbol{\lambda}) = \sum_{m=0}^{T-j-p-k-1} s_m \gamma_{m+j}^*(\boldsymbol{\lambda}), \qquad j = 1, \ldots, p$$

$$L_{p+j}^*(\boldsymbol{\lambda}) = \sum_{m=0}^{T-j-p-k-1} t_m \gamma_{m+j}^*(\boldsymbol{\lambda}), \qquad j = 1, \ldots, q$$

$$L_{p+q+1}^*(\boldsymbol{\lambda}) = \sum_{m=p+1+k}^{T} \psi(r_{m,k}/\hat{\sigma}_k), \tag{3.11}$$

with

$$\gamma_i^*(\boldsymbol{\lambda}) = \sum_{t=p+i+k+1}^{T} \eta(r_{t,h(i)}/\hat{\sigma}_{h(i)}, r_{t-i,k}/\hat{\sigma}_k), \tag{3.12}$$

and where $\hat{\sigma}_i$ is a scale estimate of the $r_{t,i}$'s $(k - q \leq i \leq q)$. The scale estimates $\hat{\sigma}_i$ may be obtained simultaneously with $\hat{\boldsymbol{\lambda}}$.

It may be proved that the TRA estimates are under general conditions asymptotically normal. A heuristic proof and the asymptotic covariance matrix expression are given in Appendix A. The asymptotic covariance matrix is very complicated and it seems difficult to compute.

We conjecture that if η and ψ are continuous and bounded, the TRA estimates are qualitative robust and resistant in the sense of strong pointwise robustness defined in Boente et al. (1982). We could not give a formal proof of this fact due to the same reasons mentioned in Section 2.5 for the RA estimates.

In Section 4 we provide results of a Monte Carlo study of the behavior of TRA estimates for the MA(1) model.

4. MONTE CARLO RESULTS

4.1 The Procedure

In this section we exhibit the results of a Monte Carlo study comparing the LS, M, GM, and TRA estimates for Gaussian,

Table 4. Monte Carlo Results for the MA(1) Process With $\theta = -.5$

	$\varepsilon = 0$			$\varepsilon = .05$, $\tau^2 = 9 \text{ var } w_t$			$\varepsilon = .05$, $\tau^2 = 100 \text{ var } w_t$		
Estimates	Mean	MSE	Eff	Mean	MSE	Eff	Mean	MSE	Eff
LS	−.494	.83	1.00	−.316	5.07	1.00	−.101	17.43	1.00
*MH*95	−.501	.88	.94	−.328	4.39	1.16	−.112	16.32	1.07
RA*HH*95	−.500	.87	.95	−.380	2.58	1.97	−.267	6.63	2.63
2TRA*HH*95	−.501	.98	.85	−.386	2.58	1.97	−.283	6.42	2.71
3TRA*HH*95	−.499	.87	.95	−.381	2.57	1.97	−.270	6.59	2.65
4TRA*HH*95	−.500	.87	.95	−.380	2.58	1.97	−.268	6.63	2.63
RA*MH*95	−.501	.88	.94	−.382	2.45	2.07	−.269	6.44	2.71
2TRA*MH*95	−.501	.97	.86	−.386	2.43	2.09	−.278	6.21	2.81
3TRA*MH*95	−.499	.85	.98	−.383	2.45	2.07	−.270	6.41	2.72
4TRA*MH*95	−.501	.87	.95	−.382	2.46	2.06	−.269	6.44	2.71
*MB*95	−.501	.86	.97	−.333	4.29	1.18	−.119	15.92	1.10
RA*HB*95	−.499	.87	.95	−.412	1.92	2.64	−.355	3.26	5.34
2TRA*HB*95	−.499	.93	.89	−.410	2.18	2.33	−.413	3.49	5.00
3TRA*HB*95	−.499	.84	.99	−.403	2.11	2.40	−.379	3.35	5.20
4TRA*HB*95	−.499	.84	.99	−.402	2.10	2.41	−.362	3.36	5.19
RA*MB*95	−.500	.85	.98	−.416	1.70	2.98	−.372	2.53	6.90
2TRA*MB*95	−.499	.92	.90	−.415	1.88	2.70	−.419	2.25	7.76
3TRA*MB*95	−.499	.85	.98	−.408	1.85	2.74	−.386	2.48	7.04
4TRA*MB*95	−.500	.84	.99	−.406	1.87	2.71	−.376	2.54	6.85
RA*HB*90	−.500	.92	.90	−.411	1.89	2.68	−.374	2.76	6.32
2TRA*HB*90	−.502	1.05	.79	−.424	1.93	2.63	−.431	2.96	5.89
3TRA*HB*90	−.500	.93	.89	−.415	1.91	2.65	−.403	3.05	5.71
4TRA*HB*90	−.500	.92	.90	−.412	1.91	2.65	−.382	2.86	6.10
RA*MB*90	−.501	.94	.88	−.420	1.60	3.17	−.395	2.02	8.64
2TRA*MB*90	−.502	1.04	.80	−.432	1.66	3.05	−.446	1.91	9.13
3TRA*MB*90	−.500	.94	.88	−.424	1.62	3.13	−.416	2.03	8.60
4TRA*MB*90	−.500	.93	.89	−.422	1.61	3.15	−.401	2.07	8.42
RA*HB*80	−.501	1.07	.78	−.425	1.69	3.00	−.398	2.18	8.01
RA*MB*80	−.501	1.10	.76	−.434	1.49	3.40	−.415	1.71	10.20
RA*HB*70	−.501	1.24	.67	−.434	1.75	2.89	−.413	1.98	8.80
RA*MB*70	−.501	1.31	.64	−.441	1.54	3.30	−.426	1.64	10.60
RA*HB*60	−.499	1.39	.60	−.440	1.89	2.68	−.420	1.96	8.90
RA*MB*60	−.500	1.49	.56	−.445	1.66	3.05	−.432	1.73	10.10

NOTE: Sample size—100; 500 replications. See Appendix C for an explanation of the notation.

MA(1), and AR(1) models without outliers $v_t \equiv 0$ in (1.9) and with additive outliers [$v_t \not\equiv 0$ in (1.9)].

For the AR(1) model the w_t in (1.9) are given by

$$w_t = \phi w_{t-1} + u_t, \qquad 1 \le t \le T, \tag{4.1}$$

and for the MA(1) model,

$$w_t = u_t - \theta u_{t-1}, \qquad 1 \le t \le T, \tag{4.2}$$

where the u_t are iid random variables $N(0, 1)$. The v_t in (1.9) are iid random variables that are independent of u_t and have distribution (1.10) with

$$G = N(0, \tau^2). \tag{4.3}$$

The original purely Gaussian case corresponds to $\varepsilon = 0$.

For the various estimates we obtain the following:

1. An estimate of their asymptotic values (and hence asymptotic bias) under particular additive outliers models, computed using two samples of size 10,000.

2. Monte Carlo estimates of their mean squared error (MSE) for samples of size 100 both with and without additive outliers. The number of replications in this case is 500. For the AR(1) process with additive outliers we use $\varepsilon = .1$ and $\tau^2 = 9$ var w_t. For the MA(1) process we use the additive outlier models (a) $\varepsilon = .05$, $\tau^2 = 9$ var w_t; (b) $\varepsilon = .05$, $\tau^2 = 100$ var w_t; and (c) $\varepsilon = .1$, $\tau^2 = 9$ var w_t. Since for model (c) the asymptotic values obtained according to estimate 1 presented too large biases, we have not performed the Monte Carlo study for sample size 100.

The values used of ϕ and θ are $\pm.5$ and $\pm.8$. Since the results for negative values are similar to those for positive values, we report here only the AR(1) cases of $\phi = .5$ and .8 and the MA(1) cases of $\theta = -.5$ and $-.8$ (the latter yield positive lag-one correlations).

We use two types of functions—Hampel and Mallows, given by (2.9′). We also use two families of ψ functions—the Huber family $\psi_{H,c}$ defined by (2.10) and the bisquare family $\psi_{B,c}$ defined by (2.11). In both cases c is the tuning constant.

Combining the two families of ψ functions with the two types of η functions, we obtain four families of η functions—$\eta_{HH,c}$, $\eta_{HB,c}$, $\eta_{MH,c}$, $\eta_{MB,c}$, where the first subscript denotes the type of η functions, the second the family of ψ functions, and c is the tuning constant of the corresponding ψ functions.

In this study we consider LS estimates, defined in Section 1, and M estimates. The latter are defined in Remark 1 of Section 2. They are asymptotically equivalent to the RA estimates with $\eta(u, v) = \psi(u)v$, where $\psi(u) = \rho'(u)$. We consider two M estimates—MH, based on $\psi_{H,c}$, and MB, based on $\psi_{B,c}$.

3. RA estimates: they are defined for the MA(1) model by (2.14) and (2.15) with $\mu = 0$. For the AR(1) model they are defined by (2.16) and (2.17) with $\mu = 0$. We consider four different types of RA estimates—RA*HH*, based on $\eta_{HH,c}$; RA*HB*, based on $\eta_{HB,c}$; RA*MH*, based on $\eta_{MH,c}$; and RA*MB*, based on $\eta_{MB,c}$.

Table 5. Monte Carlo Results for the MA(1) Process With $\theta = -.8$

	$\varepsilon = 0$			$\varepsilon = .05, \tau^2 = 9$ var w_t			$\varepsilon = .05, \tau^2 = 100$ var w_t		
Estimates	Mean	MSE	Eff	Mean	MSE	Eff	Mean	MSE	Eff
LS	−.791	.49	1.00	−.426	16.37	1.00	−.127	47.42	1.00
*MH*95	−.792	.51	.96	−.449	14.45	1.13	−.148	44.59	1.06
RA*HH*95	−.792	.51	.96	−.525	9.04	1.81	−.357	21.45	2.21
2TRA*HH*95	−.785	.79	.62	−.553	8.05	2.03	−.412	18.32	2.59
3TRA*HH*95	−.789	.65	.75	−.537	8.67	1.89	−.372	20.64	2.30
4TRA*HH*95	−.791	.56	.88	−.530	8.92	1.84	−.362	21.25	2.23
RA*MH*95	−.793	.51	.96	−.526	8.86	1.85	−.357	21.25	2.23
2TRA*MH*95	−.789	.85	.58	−.548	8.06	2.03	−.394	18.87	2.51
3TRA*MH*95	−.790	.63	.78	−.535	8.58	1.91	−.369	20.63	2.30
4TRA*MH*95	−.791	.55	.89	−.530	8.77	1.87	−.361	21.08	2.25
*MB*95	−.792	.51	.96	−.456	13.95	1.17	−.167	42.64	1.11
RA*HB*95	−.792	.51	.96	−.563	7.01	2.34	−.478	12.01	3.95
2TRA*HB*95	−.787	.82	.60	−.614	5.81	2.82	−.611	7.52	6.30
3TRA*HB*95	−.790	.66	.74	−.593	6.40	2.56	−.566	9.37	5.06
4TRA*HB*95	−.790	.56	.88	−.577	6.78	2.41	−.525	10.92	4.34
RA*MB*95	−.792	.51	.96	−.563	6.75	2.43	−.495	10.46	4.53
2TRA*MB*95	−.788	.84	.58	−.605	5.59	2.93	−.617	5.68	8.35
3TRA*MB*95	−.788	.63	.78	−.581	6.23	2.63	−.572	7.73	6.13
4TRA*MB*95	−.791	.56	.88	−.573	6.54	2.50	−.531	9.37	5.06
RA*HB*90	−.792	.52	.94	−.581	6.09	2.69	−.505	10.27	4.62
2TRA*HB*90	−.788	.87	.56	−.636	4.93	3.32	−.645	5.93	8.00
3TRA*HB*90	−.789	.70	.70	−.612	5.52	2.97	−.602	7.52	6.30
4TRA*HB*90	−.791	.58	.85	−.598	5.81	2.82	−.556	9.05	5.24
RA*MB*90	−.792	.54	.91	−.589	5.49	2.98	−.531	8.33	5.69
2TRA*MB*90	−.787	.89	.55	−.639	4.32	3.79	−.652	4.28	11.10
3TRA*MB*90	−.788	.66	.74	−.619	4.80	3.41	−.624	5.30	8.95
4TRA*MB*90	−.791	.59	.83	−.603	5.20	3.15	−.584	6.81	6.97
RA*HB*80	−.791	.62	.80	−.604	5.09	3.22	−.540	8.10	5.85
RA*MB*80	−.791	.65	.76	−.615	4.45	3.68	−.566	6.50	7.30
RA*HB*70	−.789	.74	.67	−.620	4.45	3.68	−.562	6.94	6.84
RA*MB*70	−.789	.78	.64	−.630	3.95	4.14	−.586	5.60	8.46
RA*HB*60	−.788	.85	.58	−.630	4.10	3.99	−.577	6.23	7.61
RA*MB*60	−.787	.90	.54	−.640	3.69	4.44	−.599	5.12	9.26

NOTE: Sample size—100; 500 replications. See Appendix C for an explanation of the notation.

4. TRA estimates: they are studied only for the MA(1) model, and defined by (3.1), (3.1′) (with $\gamma_{j,k}$ instead of γ_j), (3.2), and (3.3). We consider the TRA estimates k-TRA*HH*, based on $\eta_{HH,c}$; k-TRA*HB*, based on $\eta_{HB,c}$; k-TRA*MH*, based on $\eta_{MH,c}$; and k-TRA*MB*, based on $\eta_{MB,c}$. k is the truncation parameter and is taken equal to 2, 3, and 4.

5. GM estimates: they are proposed in Denby and Martin (1979). They are defined by

$$\sum_{t=2}^{T} \eta(r_t/\hat{\sigma}_r, z_{t-1}/\hat{\sigma}_z) = 0,$$

with

$$\hat{\sigma}_r = \text{Med}(|r_2|, \ldots, |r_T|)/.6745$$

and

$$\hat{\sigma}_z = \text{Med}(|z_1|, \ldots, |z_T|)/.6745.$$

They are only studied for the AR(1) model. We consider four different types of GM estimates—GM*HH*, based on $\eta_{HH,c}$; GM*HB*, based on $\eta_{HB,c}$; GM*MH*, based on $\eta_{MH,c}$; and GM*MB* based on $\eta_{MB,c}$.

In the case of the AR(1) model, all the constants of the M, GM, and RA estimates are tuned so that the asymptotic relative efficiency with respect to the LS estimate under a purely Gaussian model (AREG) is .95. RA and GM estimates based on the same η functions have the same AREG. Therefore we use the same constants for both classes of estimates.

In the case of the MA(1) model, we consider M and RA estimates based on $\psi_{H,c}$ with AREG .95. The RA estimates based on $\psi_{B,c}$ considered in this study are those with AREG .95, .90, .80, .70, and .60. The k-TRA estimates considered here have the same constants that the corresponding RA estimates with AREG .95 and .90 have for those based on $\psi_{B,c}$. As expected, however, the Monte Carlo results show that these k-TRA estimates have smaller AREG than the corresponding RA estimates. The Monte Carlo also shows that the AREG of the TRA estimates increases with k and decreases with $|\theta|$, also as expected.

In Table 1 we show the constants of the RA and M estimates used in this study.

We identify each M, GM, and RA estimate by writing the AREG times 100 after its name; for example, RA*HB*90 denotes the RA*HB* estimate with AREG .90. For the TRA estimates we use the same notation, except with a numerical prefix to indicate the value of k, but using the AREG of the RA estimate with the same η function.

We use certain routines from the International Mathematical and Statistical Libraries (IMSL, 1982) in FORTRAN. All remaining computer programs are written in FORTRAN and they

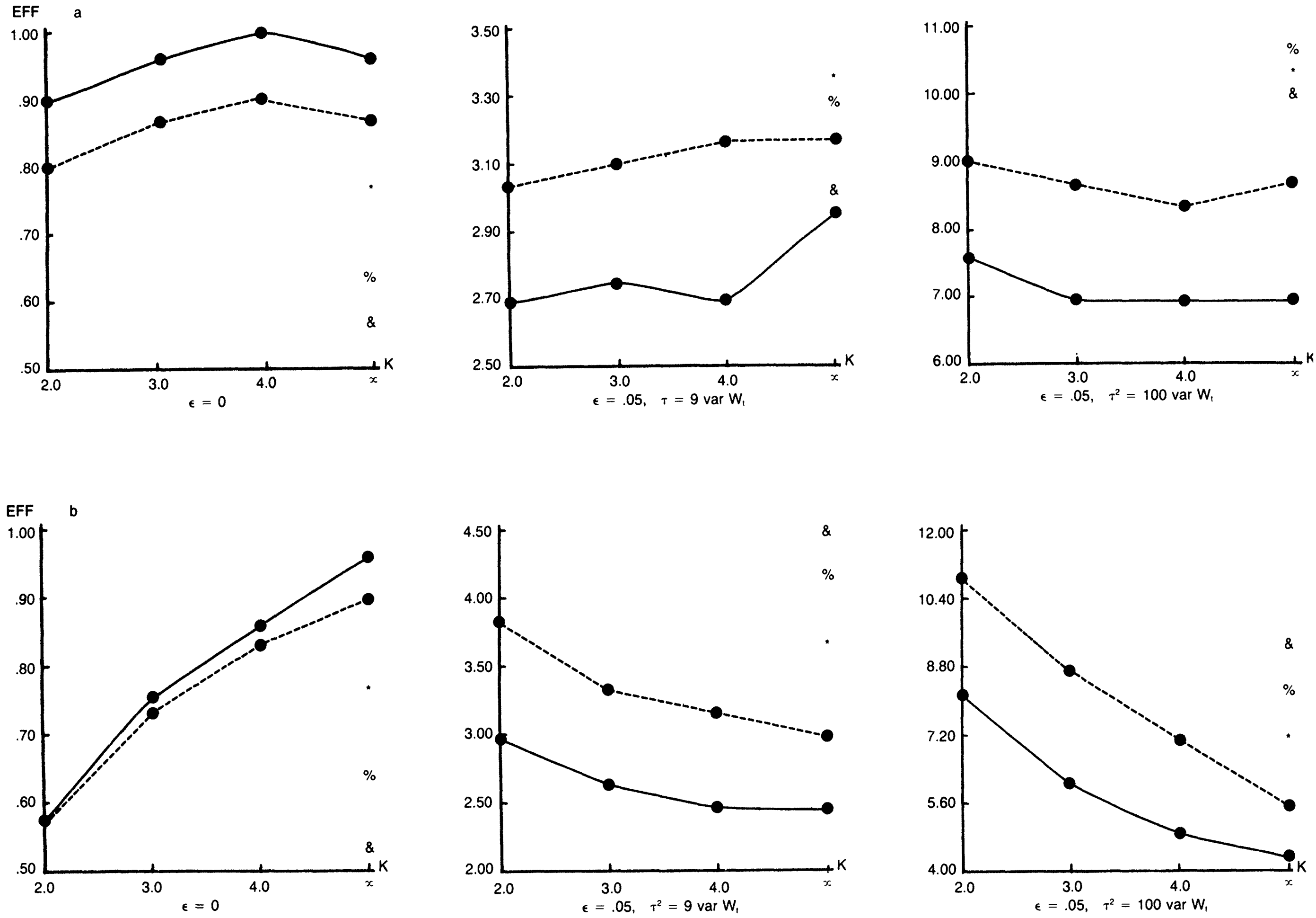

*Figure 2. Efficiency of the RA and TRA Mallows-Type Estimates, Based on $\psi_{B,c}$ for the Pure and Additive Outlier MA(1) Models: sample size 100; ——, TRAMB95; ---, TRAMB90; *, RAMB80; %, RAMB70; &, RAMB60; K, truncation parameter; K = ∞ corresponds to an RA estimate. a—θ = .5; b—θ = .8.*

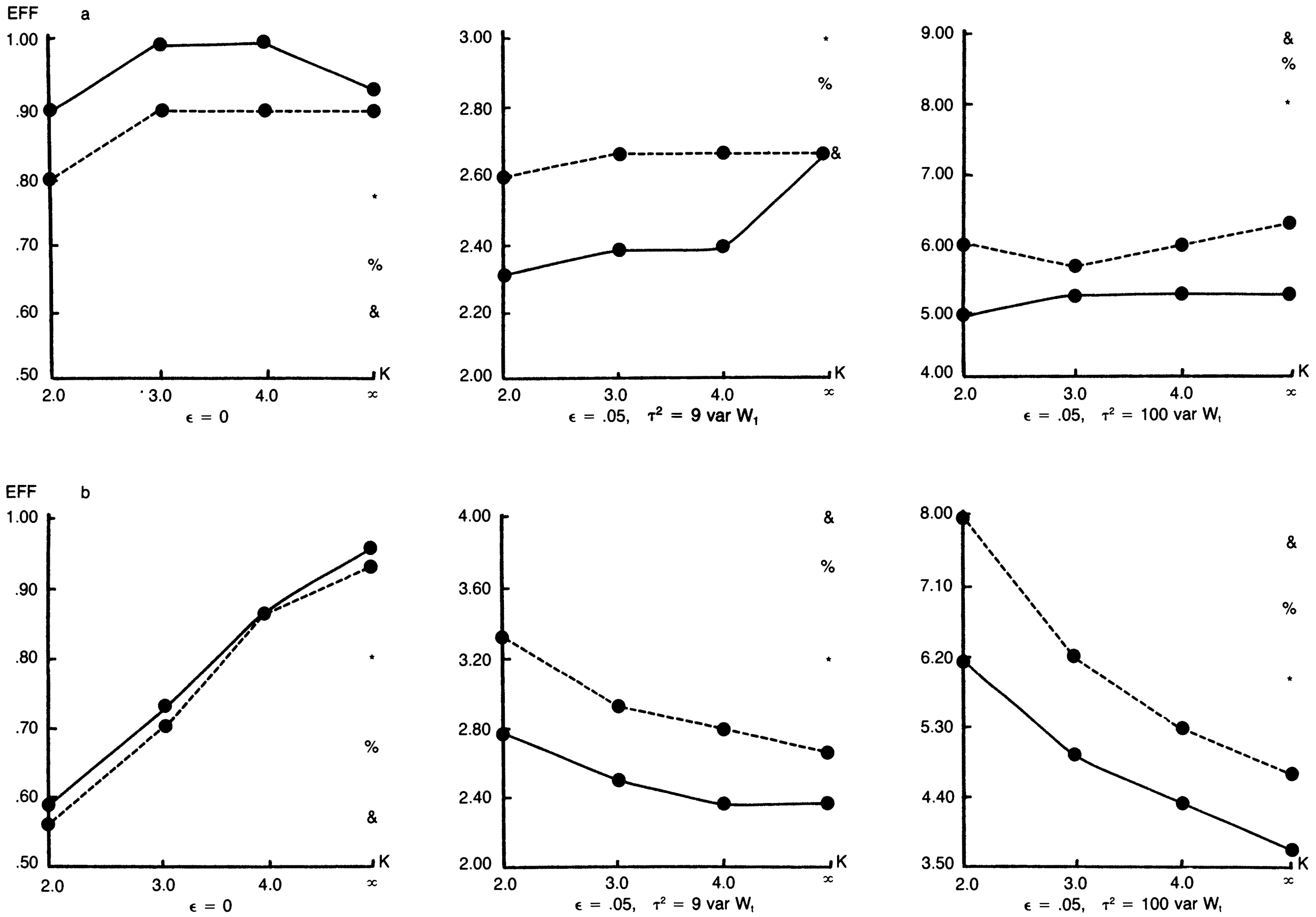

*Figure 3. Efficiency of the RA and TRA Hampel-Type Estimates, Based on $\psi_{B,c}$ for the Pure and Additive Outlier MA(1) Models: sample size 100; ——, TRAHB95; ---, TRAHB90; *, RAHB80; %, RAHB70; &, RAHB60; K, truncation parameter; K = ∞ corresponds to an RA estimate. a—$\theta = .5$; b—$\theta = .8$.*

are available upon request. We ran all of the computer programs on the IBM 370/158 computer system [Michigan Terminal System (MTS) operating system] at the Instituto de Matemática Pura e Aplicada (IMPA). We give some details concerning random number generation and estimate calculations in Appendix B. We made the several tables using some BMDP (1979) programs.

4.2 Results for the AR(1) Model

Table 2 shows the results of the Monte Carlo study with 500 replications of samples of size 100 for the purely AR(1) Gaussian model (ARG model) and for an AR(1) Gaussian model with additive outliers (ARAO) with $\varepsilon = .1$ and $\tau^2 = 9$ var w_t. In Figure 1 we plot the efficiency of the estimates for the two models.

Table 3 shows the estimates of the asymptotic values, the square bias, and the relative bias of the different estimates under the ARAO model with $\varepsilon = .1$ and $\tau^2 = 9$ var w_t.

Under the ARG model the MSE's are greater than the asymptotic variances. For example, for $\phi = .5$ the LS estimate has MSE .85 and the asymptotic variance is .75. For $\phi = .8$ the MSE is .51 and the asymptotic variance is .36.

Under the ARAO model, the RA and GM estimates perform better than the LS and M estimates. The estimates based on $\psi_{B,c}$ perform better than those based on $\psi_{H,c}$. The RA and GM estimates behave similarly for $\phi = .5$, but the RA estimates are superior for $\phi = .8$.

For $\phi = .5$ the asymptotic values based on Hampel- and Mallows-type η functions are similar, but the MSE for samples of size 100 is smaller when the Mallows-type η functions are used. For $\phi = .8$ the estimates based on the Hampel-type η functions are better with regard to both asymptotic values and MSE for samples of size 100.

4.3 Results for the MA(1) Model

In Tables 4 and 5 we give the results of the Monte Carlo study using 500 replications corresponding to (a) a purely Gaussian MA(1) model (MAG model), (b) an MA(1) Gaussian model with additive outliers (MAAO model) with $\varepsilon = .05$ and $\tau^2 = 9$ var w_t (MAAO1 model), and (c) an MAAO model with $\varepsilon = .05$ and $\tau^2 = 100$ var w_t (MAAO2 model).

In Figure 2 we plot the efficiency of the Mallows-type RA and TRA estimates for the same three models. Similar plots for the Hampel RA and TRA estimates are given in Figure 3.

In Tables 6 and 7 we give the asymptotic results for an MAAO model with (a) $\varepsilon = .05$, $\tau^2 = 9$ var w_t (MAAO1 model); (b) $\varepsilon = .05$, $\tau^2 = 100$ var w_t (MAAO2 model); and (c) $\varepsilon = .1$, $\tau^2 = 9$ var w_t (MAAO3 model).

For the MAG model we observe again that the MSE's of LS, M, and RA estimates are larger than their asymptotic variances. We also observe for this model that the k-TRA estimates ($k = 2, 3, 4$) have a larger MSE than the corresponding RA estimates when $\theta = -.8$. If $\theta = -.5$, the 2-TRA estimates have a larger MSE than the RA estimates, but for $k = 3, 4$, both classes of estimates behave similarly.

The asymptotic values for the MAAO model with $\varepsilon = .1$ and $\tau^2 = 9 \operatorname{var} w_t$ show that all of the estimates have a relatively large asymptotic bias. This is because one outlier z_t has influence on all of the residuals r_r with $r \geq t$ and on the truncated residuals, $r_{t+i,k}$'s ($0 \leq i \leq k$). Therefore 10% of outliers in the z_t's implies a larger percentage of outliers in the r_t's and in the $r_{t,k}$'s. This percentage will increase with θ. We observe, however, that the RA and TRA estimates have smaller asymptotic biases than the LS and M estimates.

Estimates based on the $\psi_{B,c}$ function have smaller asymptotic biases than those based on the $\psi_{H,c}$ function. Since all of the considered estimates have large asymptotic biases, and since we could afford only a limited computing time, we preferred to perform the Monte Carlo study for samples of size 100 with $\varepsilon = .05$.

For the MAAO model with $\varepsilon = .05$, we observe that in both cases $\tau^2 = 9 \operatorname{var} w_t$ and $\tau^2 = 100 \operatorname{var} w_t$, the RA and TRA estimates performed much better than the LS and M estimates. Once more, estimates based on a $\psi_{B,c}$ function are better than the corresponding estimates based on a $\psi_{H,c}$ function. The above comparisons hold for the square and relative biases when the sample size is 100 and for asymptotic values. For samples of size 100 the relative bias of estimates of the Mallows type are smaller than the corresponding estimates of the Hampel type. The asymptotic values for estimates of the Mallows type are similar to the corresponding estimates of the Hampel type, except for the MAAO3 model, where the former are slightly better than the latter.

We observe that the biases under the MAAO models are especially large for $\theta = -.8$. The corresponding Figures 2 and 3 show that it is precisely for this value that the TRA estimates performed better than the corresponding RA estimates. As expected, the improvement of the TRA estimates with respect to the RA estimates is greater for the large-variance gross-error MAAO model corresponding to $\tau^2 = 100 \operatorname{var} w_t$. For $\theta = -.5$ the performance of RA and TRA estimates under the MAAO models is similar. At the same time, for these values of θ the loss in efficiency under the MAG model of the TRA estimates with respect to the RA estimates is small or null. On the other hand, an RA estimate with high efficiency under the MAAO models for $\theta = -.8$—for example, RA*MB*60 or RA*HB*60—has low efficiency under the MAG model for all θ. This is related to the fact that the AREG of the RA estimates is independent of θ.

We can also observe in Figures 2 and 3 that the TRA*MB*90 and TRA*HB*90 estimates seem to be preferable to the TRA*MB*95 and TRA*HB*95 estimates. Indeed, for $\theta = -.8$ and for the

Table 6. Asymptotic Results for the MAAO Models: $\theta = -.5$

	$\varepsilon = .05$, $\tau^2 = 9 \operatorname{var} w_t$			$\varepsilon = .05$, $\tau^2 = 100 \operatorname{var} w_t$			$\varepsilon = .1$, $\tau^2 = 9 \operatorname{var} w_t$		
Estimates	AV	Bias 2	RELB	AV	Bias 2	RELB	AV	Bias 2	RELB
LS	−.30	.040	.40	−.06	.194	.88	−.22	.078	.56
*MH*95	−.32	.032	.36	−.07	.185	.86	−.23	.073	.54
RA*HH*95	−.38	.014	.24	−.27	.053	.46	−.30	.040	.40
2TRA*HH*95	−.39	.012	.22	−.28	.048	.44	−.30	.040	.40
3TRA*HH*95	−.38	.014	.24	−.27	.053	.46	−.30	.040	.40
4TRA*HH*95	−.38	.014	.24	−.27	.053	.46	−.30	.040	.40
RA*MH*95	−.38	.014	.24	−.27	.053	.46	−.31	.036	.38
2TRA*MH*95	−.39	.012	.22	−.28	.048	.44	−.31	.036	.38
3TRA*MH*95	−.39	.012	.22	−.27	.053	.46	−.31	.036	.38
4TRA*MH*95	−.38	.014	.24	−.27	.053	.46	−.31	.036	.38
*MB*95	−.33	.029	.34	−.07	.185	.86	−.22	.078	.56
RA*HB*95	−.41	.008	.18	−.37	.017	.26	−.33	.029	.34
2TRA*HB*95	−.41	.008	.18	−.42	.006	.16	−.33	.029	.34
3TRA*HB*95	−.41	.008	.18	−.37	.017	.26	−.33	.029	.34
4TRA*HB*95	−.41	.008	.19	−.37	.017	.26	−.33	.029	.34
RA*MB*95	−.41	.008	.18	−.38	.014	.24	−.34	.026	.32
2TRA*MB*95	−.42	.006	.16	−.42	.006	.16	−.35	.023	.30
3TRA*MB*95	−.41	.008	.18	−.39	.012	.22	−.34	.026	.32
4TRA*MB*95	−.41	.008	.18	−.38	.014	.24	−.34	.026	.32
RA*HB*90	−.42	.006	.16	−.38	.014	.24	−.35	.023	.30
2TRA*HB*90	−.43	.005	.14	−.38	.014	.24	−.35	.023	.30
3TRA*HB*90	−.42	.006	.16	−.44	.004	.12	−.35	.023	.30
4TRA*HB*90	−.42	.006	.16	−.39	.012	.22	−.35	.023	.30
RA*MB*90	−.43	.005	.14	−.38	.014	.24	−.36	.020	.28
2TRA*MB*90	−.43	.005	.14	−.40	.010	.20	−.37	.017	.26
3TRA*MB*90	−.43	.005	.14	−.45	.003	.10	−.36	.020	.28
4TRA*MB*90	−.43	.005	.14	−.42	.006	.16	−.36	.020	.28
RA*HB*80	−.43	.005	.14	−.41	.008	.18	−.37	.017	.26
RA*MB*80	−.44	.004	.12	−.40	.010	.20	−.38	.014	.24
RA*HB*70	−.44	.004	.12	−.42	.006	.16	−.38	.014	.24
RA*MB*70	−.45	.003	.10	−.42	.006	.16	−.39	.012	.22
RA*HB*60	−.45	.003	.10	−.43	.005	.14	−.39	.012	.22
RA*MB*60	−.45	.003	.10	−.43	.005	.14	−.40	.010	.20

NOTE: See Appendix C for an explanation of the notation.

Table 7. Asymptotic Results for the MAAO Models: $\theta = -.8$

	$\varepsilon = .05, \tau^2 = 9 \text{ var } w_t$			$\varepsilon = .05, \tau^2 = 100 \text{ var } w_t$			$\varepsilon = .1, \tau^2 = 9 \text{ var } w_t$		
Estimates	AV	Bias 2	RELB	AV	Bias 2	RELB	AV	Bias 2	RELB
LS	−.40	.160	.50	−.08	.518	.90	−.29	.260	.64
*MH*95	−.45	.123	.44	−.08	.518	.90	−.30	.250	.63
RA*HH*95	−.54	.068	.33	−.10	.490	.88	−.41	.152	.49
2TRA*HH*95	−.57	.053	.29	−.36	.194	.55	−.42	.144	.48
3TRA*HH*95	−.55	.063	.31	−.41	.152	.49	−.41	.152	.49
4TRA*HH*95	−.54	.068	.33	−.37	.185	.54	−.41	.152	.49
RA*MH*95	−.54	.068	.33	−.36	.194	.55	−.41	.152	.49
2TRA*MH*95	−.56	.058	.30	−.36	.194	.55	−.42	.144	.48
3TRA*MH*95	−.55	.063	.31	−.40	.160	.50	−.41	.152	.49
4TRA*MH*95	−.54	.068	.33	−.37	.185	.54	−.41	.152	.49
*MB*95	−.46	.116	.43	−.11	.476	.86	−.31	.240	.61
RA*HB*95	−.58	.048	.28	−.49	.096	.39	−.45	.123	.44
2TRA*HB*95	−.64	.026	.20	−.72	.006	.10	−.48	.102	.40
3TRA*HB*95	−.60	.040	.25	−.66	.020	.18	−.46	.116	.43
4TRA*HB*95	−.58	.048	.28	−.52	.078	.35	−.45	.123	.44
RA*MB*95	−.57	.053	.29	−.50	.090	.38	−.46	.116	.43
2TRA*MB*95	−.61	.036	.24	−.68	.014	.15	−.48	.102	.40
3TRA*MB*95	−.59	.044	.26	−.60	.040	.25	−.46	.116	.43
4TRA*MB*95	−.58	.048	.28	−.53	.073	.34	−.45	.123	.44
RA*HB*90	−.60	.040	.25	−.52	.078	.35	−.48	.102	.40
2TRA*HB*90	−.66	.020	.18	−.74	.004	.08	−.52	.078	.35
3TRA*HB*90	−.63	.029	.21	−.71	.008	.11	−.49	.096	.39
4TRA*HB*90	−.61	.036	.24	−.59	.044	.26	−.48	.102	.40
RA*MB*90	−.60	.040	.25	−.54	.068	.33	−.49	.096	.39
2TRA*MB*90	−.65	.023	.19	−.72	.006	.10	−.52	.078	.35
3TRA*MB*90	−.63	.029	.21	−.67	.017	.16	−.50	.090	.38
4TRA*MB*90	−.61	.036	.24	−.60	.040	.25	−.49	.096	.39
RA*HB*80	−.62	.032	.23	−.56	.058	.30	−.51	.084	.36
RA*MB*80	−.63	.029	.21	−.58	.048	.28	−.52	.078	.35
RA*HB*70	−.64	.026	.20	−.59	.044	.26	−.53	.073	.34
RA*MB*70	−.65	.023	.19	−.60	.040	.25	−.55	.063	.31
RA*HB*60	−.65	.023	.19	−.60	.040	.25	−.54	.068	.33
RA*MB*60	−.66	.020	.18	−.60	.040	.25	−.56	.058	.30

NOTE: See Appendix C for an explanation of the notation.

three models considered here, the 3TRA*HB*90 and the 3TRA*MB*90 estimates are much better than the 2TRA*HB*95 and 2TRA*MB*95 estimates, respectively, and the 4TRA*HB*90 and 4TRA*MB*90 estimates are much better than the 3TRA*HB*95 and 3TRA*MB*95 estimates. For $\theta = -.5$ the same comparisons show that the TRA*MB*90 and TRA*HB*90 estimates have small or null losses of efficiency under the MAG model with respect to the TRA*MB*95 and TRA*HB*95 estimates.

4.4 Conclusions

The above Monte Carlo comparisons lead us to conclude that, at least for additive outliers, the RA estimates based on the $\psi_{B,c}$ functions have very good robustness properties for the AR(1) model and they compare favorably with the GM estimates. The RA estimates also behave robustly in terms of efficiency for the MA(1) model (even if they are not qualitatively robust) when $\varepsilon \le .05$, $\tau^2 \le 100$ var w_t, and $|\theta| \le .5$. If there exists the possibility of $|\theta|$ close to 1, say larger than .8 and $\tau^2 \ge 100$ var w_t, it seems preferable to use the TRA*MB*90 or the TRA*HB*90 estimates, even if this implies some loss of efficiency under the MAG model.

Further research is needed to determine the behavior of these estimates for higher-order models and for other types of outliers—for example, in the case when there are patches of outliers. It will also be important to determine the performance of the algorithm proposed in this article and to look for more efficient alternatives.

APPENDIX A: HEURISTIC DERIVATION OF THE ASYMPTOTIC DISTRIBUTION OF RA AND TRA ESTIMATES

Let $L_j^*(\boldsymbol{\lambda})$, $1 \le j \le p + q + 1$, be as defined in (3.11). Then it is immediate that we can write

$$L_j^*(\boldsymbol{\lambda}) \equiv \sum_{t=j+k+p+1}^{T} \delta_{j,t}^*(\boldsymbol{\lambda}), \qquad 1 \le j \le p + q + 1, \tag{A.1}$$

where

$$\delta_{j,t}^*(\boldsymbol{\lambda}) = \sum_{m=0}^{t-j-p-1} s_m \eta(r_{t,h(j+m)}/\hat{\sigma}_{h(j+m)}, r_{t-j-m,k}/\hat{\sigma}_k), \quad 1 \le j \le p$$

$$\delta_{p+j,t}^*(\boldsymbol{\lambda}) = \sum_{m=0}^{t-j-p-k-1} t_m \eta(r_{t,h(j+m)}/\hat{\sigma}_{h(j+m)}, r_{t-j-m,k}/\hat{\sigma}_k), \quad 1 \le j \le q$$

$$\delta_{p+q+1}^*(\boldsymbol{\lambda}) = \psi(r_{t,k}/\hat{\sigma}_k). \tag{A.2}$$

Let $\boldsymbol{\lambda}_0$ be the true value of $\boldsymbol{\lambda}$, and let $\hat{\boldsymbol{\lambda}}$ be the TRA-estimate solution of (3.10). Then using a Taylor expansion of (3.10), we get asymptotically

$$T^{1/2}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0) \cong -(DL^*(\boldsymbol{\lambda}_0)/T)^{-1}(L^*(\boldsymbol{\lambda}_0)/T^{1/2}), \tag{A.3}$$

where $D\mathbf{L}^*(\boldsymbol{\lambda}_0)$ is the differential matrix of $\mathbf{L}^*$ with respect to $\boldsymbol{\lambda}$.

We will give heuristic proofs of

$$L^*(\boldsymbol{\lambda}_0)/T^{1/2} \xrightarrow{d} N(\mathbf{0}, \mathbf{A}) \tag{A.4}$$

and

$$DL^*(\boldsymbol{\lambda}_0)/T \xrightarrow{p} \mathbf{B}, \tag{A.5}$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\xrightarrow{p}$ denotes convergence in probability and where **A** and **B** are $(p + q + 1) \times (p + q + 1)$ matrices that will be defined in (A.9) and (A.11), respectively. Therefore we have

$$T^{1/2}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1\prime}). \tag{A.6}$$

Assume that $\hat{\sigma}_k \xrightarrow{p} \sigma_k$ $(k - q \le j \le k)$. Then, since the sequences s_i and t_i decay exponentially when $i \to \infty$, we have asymptotically

$$L_j^*(\boldsymbol{\lambda}) \cong \sum_{t=1}^{\infty} \delta_{j,t}(\boldsymbol{\lambda}), \qquad 1 \le j \le p + q + 1, \tag{A.7}$$

where

$$\delta_{j,t}(\boldsymbol{\lambda}) = \sum_{m=0}^{\infty} s_m \eta(r_{t,h(j+m)}/\sigma_{h(j+m)}, r_{t-j-m,k}/\sigma_k), \quad 1 \le j \le p$$

$$\delta_{p+j,t}(\boldsymbol{\lambda}) = \sum_{m=0}^{\infty} t_m \eta(r_{t,h(j+m)}/\sigma_{h(j+m)}, r_{t-j-m,k}/\sigma_k), \quad 1 \le j \le q$$

$$\delta_{p+q+1,t}(\boldsymbol{\lambda}) = \psi(r_{t,k}/\sigma_k). \tag{A.8}$$

Since the multivariate process $\boldsymbol{\delta}_t(\boldsymbol{\lambda}_0) = (\delta_{1,t}(\boldsymbol{\lambda}_0), \ldots, \delta_{p+q+1,t}(\boldsymbol{\lambda}_0))$ is stationary and ergodic, (A.7) and the central limit theorem will imply (A.4), where the i, j element of **A** is given by

$$A_{i,j} = \sum_{m=-\infty}^{\infty} v_{i,j}(m), \tag{A.9}$$

where

$$v_{i,j}(m) = E(\delta_{i,t}(\boldsymbol{\lambda}_0)\delta_{j,t-m}(\boldsymbol{\lambda}_0)). \tag{A.10}$$

On the other hand, by the ergodic theorem, (A.5) holds with

$$\mathbf{B} = E(D\boldsymbol{\delta}_t(\boldsymbol{\lambda}_0)). \tag{A.11}$$

$v_{i,j}(m)$ may be estimated by $\hat{v}_{i,j}(m) = (\Sigma\delta_{i,t}^*(\hat{\boldsymbol{\lambda}})\delta_{j,t-m}^*(\hat{\boldsymbol{\lambda}}))/T$, and the matrix B may be estimated by $\hat{\mathbf{B}} = \Sigma_{t=p+k+1}^{T} D\boldsymbol{\delta}_t(\hat{\boldsymbol{\lambda}})/T$. Similarly, it may be shown that (A.6) holds for **RA** estimates, with **A** and **B** given by (A.9) and (A.11), but with $\delta_{j,t}(\boldsymbol{\lambda})$ given by

$$\delta_{j,t}(\boldsymbol{\lambda}) = \sum_{m=0}^{\infty} s_m \eta(r_t/\sigma, r_{t-j-m}/\sigma) \quad \text{if } 1 \le j \le p,$$

$$\delta_{p+j,t}(\boldsymbol{\lambda}) = \sum_{m=0}^{\infty} t_m \eta(r_t/\sigma, r_{t-j-m}/\sigma) \quad \text{if } 1 \le j \le q,$$

$$\delta_{p+q+1,t}(\boldsymbol{\lambda}) = \psi(r_t/\sigma).$$

We will compute explicitly the matrices **A** and **B** in this case. If $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$, then we have $r_t \cong u_t$. Take $1 \le i \le j \le p$. Therefore, since the u_t's are iid with symmetric distribution and we assume that the function $\eta(u, v)$ is odd on each variable, we have

$$v_{i,j}(m) = 0 \qquad \text{if } m \ne 0$$

$$= \left(\sum_{h=0}^{\infty} s_h s_{h+(j-i)}\right) E(\eta^2(u_t/\sigma, u_{t-1}/\sigma)) \quad \text{if } m = 0.$$

Therefore, according to (A.9), (2.19), and (2.20), we have

$$A_{i,j} = aC_{i,j}. \tag{A.12}$$

Similarly, we may prove that (A.12) holds for $1 \le i, j \le p + q$.

On the other hand, it is immediate that

$$v_{p+q+1,j}(m) = 0\ \forall m \quad \text{if } j \ne p + q + 1$$

$$= 0 \qquad \text{if } j = p + q + 1 \text{ and } m \ne 0$$

$$= a^* \qquad \text{if } j = p + q + 1 \text{ and } m = 0,$$

where a^* is defined by (2.23). Then by (A.9) we have

$$A_{i,p+q+1} = A_{p+q+1,i} = 0 \quad \text{if } i \ne p + q + 1$$

$$A_{p+q+1,p+q+1} = a^*. \tag{A.13}$$

Now we will compute the matrix **B** given by (A.11). Suppose $1 \le i \le j \le p$. Then

$$B_{i,j} = -\sum_{m=0}^{\infty} \partial s_m/\partial\phi_j E\eta(u_t/\sigma, u_{t-i-m}/\sigma)$$

$$- (1/\sigma) \sum_{m=0}^{\infty} s_m E(\eta_1(u_t/\sigma, u_{t-i-m}/\sigma)\partial u_t/\partial\phi_j)$$

$$- (1/\sigma) \sum_{m=0}^{\infty} s_m E(\eta_2(u_t/\sigma, u_{t-i-m}/\sigma)\partial u_{t-i-m}/\partial\phi_j),$$

where $\eta_1(u, v) = \partial\eta(u, v)/\partial u$ and $\eta_2(u, v) = \partial\eta(u, v)/\partial v$. The first sum is clearly equal to 0. Then using (2.2), we have

$$B_{i,j} = (1/\sigma) \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} s_m s_n E(\eta_1(u_t/\sigma, u_{t-i-m}/\sigma)u_{t-j-n})$$

$$+ (1/\sigma) \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} s_m s_n E(\eta_2(u_t/\sigma, u_{t-i-m}/\sigma)u_{t-i-m-j-n}).$$

Since $u_{t-i-m-j-n}$ is always independent of u_{t-i-m}, the second double sum is 0. The only terms in the first double sum that are different from 0 are those with $t - i - m = t - j - n$. Then we have

$$B_{i,j} = (1/\sigma) \sum_{m=0}^{\infty} s_m s_{m+(j-i)} E(\eta_1(u_t/\sigma, u_{t-1}/\sigma)u_{t-1}).$$

Then according to (2.19) and (2.21), we have for $1 \le i \le j \le p$,

$$B_{i,j} = (b/\sigma)C_{ij}. \tag{A.14}$$

Similarly, it may be shown that (A.14) holds for all i, j, $1 \le i, j \le p + q$. It is easy to show that $B_{p+q+1,j} = B_{j,p+q+1} = 0$ if $j \ne p + q + 1$.

Finally, according to (2.2) and (2.24), we have

$$B_{p+q+1,p+q+1} = E(\partial\psi(u_t/\sigma)/\partial\mu) = Hb^*/\sigma. \tag{A.15}$$

From (A.12), (A.13), (A.14), and (A.15), we have $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1} = \mathbf{D}^*$, where $\mathbf{D}^*$ is given by (2.18).

APPENDIX B: COMPUTATIONAL DETAILS

The uniform, $U(0, 1)$, numbers needed to produce both the v_t's in (1.9) and the u_t's in (4.1) and (4.2) were generated using the GGUW routine from IMSL (1982), which is a uniform (0, 1) random number generator with shuffling. Normal random numbers were generated using the Box–Knuth method. To obtain w_0 in (4.1) and (4.2), we generated an initial block of 100 u_t's and used the final number of this block as the value w_0. Once this value was obtained, we generated the z_t's in (1.9) implementing the following algorithm.

1. Generate a block of T u_t's.
2. For each $t = 1, \ldots, T$: (a) define w_t from (4.1) or (4.2); and (b) (only for AO models) draw x from GGUW if $x > \varepsilon$, then $z_t = w_t$ if $x \le \varepsilon$, and then draw v from random numbers $N(0, 1)$ and define $z_t = w_t + \tau v$.

Finally, to find a root $\hat{\lambda}$ in the equation defining each estimator, say

$$F(\hat{\lambda}) = 0, \tag{B.1}$$

we implemented the following algorithm.

1. Let λ_0 be a start "solution" of $F(\lambda) = 0$.
2. Look for an integer k between -15 and 15 such that

$$k/16 \le \lambda_0 < (k + 1)/16 \tag{B.2}$$

and

$$\operatorname{sgn} F(k/16) \ne \operatorname{sgn} F((k + 1)/16). \tag{B.3}$$

If there is not such a k, then look for a k that satisfies (B.3). If (B.3) is not true for all integers k between -15 and 15, then define $\hat{\lambda} = \lambda_0$. On the contrary, proceed with the following step.

3. Call ZBRENT [a routine excerpted from IMSL (1982)] in order to find a $\hat{\lambda}$ between $k/16$ and $(k + 1)/16$ such that $F(\hat{\lambda}) = 0$. If ZBRENT fails to find such a $\hat{\lambda}$, then define $\hat{\lambda} = \lambda_0$.

For $\hat{\lambda}$ = LS at the MA models, we used λ_0 such that

$$\lambda_0^2 + \lambda_0/r + 1 = 0, \qquad |\lambda_0| < 1,$$

where

$$r = \sum_{t=0}^{T-1} (z_{t+1} - \bar{z})(z_t - \bar{z}) \Big/ \sum_{t=0}^{T} (z_t - \bar{z})^2,$$

$$\bar{z} = (1/T) \sum_{t=1}^{T} z_t.$$

For RA and GM estimates we used λ_0 = LS, and for TRA estimates we used the corresponding RA estimate as λ_0. The algorithm implemented by ZBRENT is a combination of linear interpolation, inverse quadratic interpolation, and bisection (Brent 1971).

APPENDIX C: NOTATION USED IN TABLES 2–7

Mean denotes the sample mean of the 500 replications of the estimate.

MSE denotes 100 times the mean squared error.

Eff denotes the relative efficiency of the estimate with respect to the LS estimate—that is, the ratio between the MSE of the LS estimate and the MSE of the corresponding estimate.

Bias 2 denotes the square bias—that is, Bias 2 = $(\text{Mean} - \lambda)^2$.

RELB denotes the relative bias—that is, RELB = $|\text{Mean} - \lambda|/|\lambda|$, where $\lambda = \phi$ for AR(1) and $\lambda = \theta$ for MA(1).

AV stands for the asymptotic values of the estimates.

[*Received May 1983. Revised November 1984.*]

REFERENCES

Beaton, A. E., and Tukey, J. W. (1974), "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band Spectroscopic Data," *Technometrics*, 16, 147–185.

BMDP (1979), *Biomedical Computer Programs P-Series 1979*, Berkeley, CA: University of California Press.

Boente, G., Fraiman, R., and Yohai, V. (1982), "Qualitative Robustness for General Stochastic Processes," Technical Report 26, University of Washington, Dept. of Statistics.

Box, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis*, San Francisco: Holden-Day.

Brent, R. P. (1971), "An Algorithm With Guaranteed Convergence for Finding a Zero of a Function," *The Computer Journal*, 14, 422–425.

Bustos, O. H. (1981), "Qualitative Robustness for General Processes," Informes de Matemática Série B-002/81, Instituto de Matemática Pura e Aplicada, Brazil.

——— (1982), "General *M*-Estimates for Contaminated *p*th-Order Autoregressive Processes: Consistency and Asymptotic Normality," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 59, 491–504.

Bustos, O. H., Fraiman, R., and Yohai, V. (1984), "Asymptotic Behaviour of the Estimates Based on Residual Autocovariances for ARMA Models," Informes de Matemática Série B-019/84, Instituto de Matemática Pura e Aplicada, Brazil.

Cox, D. (1981), "Metrics on Stochastic Processes and Qualitative Robustness," Technical Report 3, University of Washington, Dept. of Statistics.

Denby, L., and Martin, R. D. (1979), "Robust Estimation of the First Order Autoregressive Parameter," *Journal of the American Statistical Association*, 74, 140–146.

Hampel, F. (1971), "A General Qualitative Definition of Robustness," *Annals of Mathematical Statistics*, 42, 1887–1896.

International Mathematical and Statistical Libraries, Inc. (1982), *IMSL Library Reference Manual* (9th ed.), Houston: Author.

Kanter, M., and Steiger, W. L. (1974), "Regression and Autoregression With Infinite Variances," *Advances in Applied Probability*, 6, 768–783.

Lee, C. H., and Martin, R. D. (1982), "*M*-Estimates for ARMA Processes," Technical Report 23, University of Washington, Dept. of Statistics.

Martin, R. D. (1980), "Robust Estimation of Autoregressive Models," in *Direction in Time Series*, eds. D. R. Brillinger and G. C. Tiao, Haywood, CA: Institute of Mathematical Statistics.

——— (1981), "Robust Methods in Time Series," in *Applied Time Series Analysis II*, ed. D. F. Findley, New York: Academic Press.

Papantoni-Kazakos, P., and Gray, R. M. (1979), "Robustness of Estimators and Stationary Observations," *Annals of Probability*, 7, 989–1002.

Tukey, J. W. (1976), "Usable Resistant/Robust Techniques of Analysis," in *Proceedings of the First ERDA Statistical Symposium, Los Alamos, New Mexico, Nov. 3–5, 1975*, eds. W. L. Nicholson and J. L. Harris, Battelle Pacific Northwest Labs.

Whittle, P. (1962), "Gaussian Estimation in Stationary Time Series," *Bulletin de l'Institut International de Statistique*, 39, 105–129.

Yohai, V. J., and Maronna, R. A. (1978), "Asymptotic Behavior of Least Square Estimates for Autoregressive Processes With Infinite Variances," *The Annals of Statistics*, 5, 554–560.